DOCUMENT RESUME

ED 170 368 TM 008 823

AUTHOR Bell, Anita I.
TITLE A Comparison of Three Equating Procedures on the Certifying Examination for Primary Care Physician's Assistants.
PUB DATE 11 Apr 79
NOTE 9p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Statistics; *Complexity Level; *Educational Trends; *Equated Scores; Item Analysis; Multiple Choice Tests; *Norms; Performance Tests; *Physicians Assistants; Postsecondary Education; Raw Scores; Standards; Test Items; *Work Sample Tests
IDENTIFIERS *Primary Care Physicians Assistants Certifying Exam; Rasch Model

ABSTRACT

An equating study was conducted on the Certifying Examination for Primary Care Physician's Assistants to compare the ability of current examinees with the standardization group and to determine if current test items are more difficult than previous items. Using 46 common items from the multiple choice section, the 1978 exam was equated to the 1976 version by four methods: Rasch, linear raw, linear short-cut, and traditional item statistics. Raw scores for the current examination, scored traditionally and equated to the 1976 test by the four methods, reflect the fact that current examinees performed better on the 46 common items than the reference examinees. Only the Rasch model however, was able to confirm that items on the current test were more difficult. This model can be used to determine whether or not examination standards should be raised as training programs improve. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED170364

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A COMPARISON OF THREE EQUATING PROCEDURES ON THE CERTIFYING EXAMINATION FOR PRIMARY CARE PHYSICIAN'S ASSISTANTS

Anita I. Bell
National Board of Medical Examiners

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Anita I. Bell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Presented at the 1979 AERA/NCME
April 11, 1979

SESSION #22.20

TM008 823

A COMPARISON OF THREE EQUATING PROCEDURES ON THE CERTIFYING EXAMINATION FOR PRIMARY CARE PHYSICIAN'S ASSISTANTS

Introduction

The Certifying Examination for Primary Care Physician's Assistants is a norm referenced examination which has been in existence for six years. The examination includes 225 multiple choice questions, about 20 Patient Management Problems, and three Clinical Skill Problems. The Patient Management Problems are paper and pencil simulations of a physician's assistant's encounter with a patient, and the Clinical Skill Problems involve physical examinations of patients. Approximately 1,650 examinees take the examination every year. The scores are standardized each year according to the performance of the reference group which is composed of graduates of formal training programs who are taking the examination for the first time. About 1,350 of the 1,650 examinees meet these qualifications.

About two years ago, a question was raised concerning the stability of the reference group. Physician's Assistants training programs have been in existence approximately 10 years. As the training programs gain experience, their curricula are changing and improving to meet newly assessed needs. Also, the educational background of the students who are entering the training programs is different from what it was six years ago, i.e., more students are entering the programs with baccalaureate degrees. It was thus felt that the students now taking the examinations are more able and/or better prepared than their predecessors.

The questions for this examination are written by a test committee composed of physicians, physician's assistants, and nurse practitioners who are all in some way associated with training programs. Having felt that the examinees they were now testing were more able, the test committee began to think that they were writing questions which were more difficult than the questions which

were written for the previous examinations. It was felt that an equating study should be conducted to try to answer these two questions and to determine the impact on the standard if the current examinees are indeed more knowledgeable.

Method

Using 46 common items, the section containing multiple-choice questions on the 1978 Physician's Assistants Examination was equated to the MCQ portion of the 1976 P.A. Exam via three methods.

Rasch Equating Procedure The two examinations were calibrated individually and difficulties were computed for each item as well as a mean ability estimate for each examinee population. The two examinations were linked through the 46 common items, and the ability scores on the current examination were shifted to be comparable to the expected ability scores of the reference test examinees had they taken the current examination.

Linear Raw Equating Procedure Angoff describes an equating procedure for groups not widely different in ability in Thorndike's 2nd edition of Educational Measurement. Common items are used to adjust for differences that may be found to exist between the groups taking the two examinations. Estimates of the mean and variance for both examinations separately are made for the combined group of all examinees by using the performance of the three groups (current exam, reference exam, and combined group) on the common items. The estimates are then used to yield a linear equation which equates the raw scores on the reference examination to the raw scores on the current examination.

A "short-cut" version of this equating procedure was also carried out. This procedure determines the ratio of standard deviations of the current

examinees and reference examinees on the common items, assumes that this ratio applies to the total test, and adjusts the current total test standard deviation to bring it in line with the reference examinees. The means of the two groups on the common items are then compared in standard deviation units and, as with the standard deviation, the current total test mean is adjusted to become an estimate for the reference examinees.

Item Statistics Equating Procedure This method uses the same formulae as the linear "short-cut", only the estimated mean and standard deviation for the reference examinees on the common items is used instead of the actual mean. This procedure is not generally used on an examination where all the common items come from one examination, but it is useful for examinations which have link items that come from more than one examination. The mean score of the reference examinees on the common items is estimated by summing the item P values, and the standard deviation is estimated from the item discrimination index as described by Harold Gulliksen in Theory of Mental Tests. These estimates are then compared with the actual mean and standard deviation of the current examinees on the same items.

Results

The raw scores for the current Physician's Assistants Examination scored traditionally and equated to the reference P.A. Examination via four methods are given in Table 1. The results of all of the methods reflect the fact that the current examinees performed better on the 46 common items than the reference examinees. Looking at the raw scores for the different methods at the mean score of 500, one can see that there is not much difference between the methods at this point on the scale. As one departs from the mean, the

Rasch procedure tends to diverge from the other methods. This is because the Rasch model results in an S-shaped curve rather than a straight line. In the area around the mean where the straight line of the other methods and the linear part of the S coincide, the scores are closely related. The further away you get from the mean, especially at the lower end of the distribution, the more divergent the sets of scores. Thus, if one were to look at 410 as a passing score, the raw score under the Rasch procedure is 4 points less than the raw score for the Linear "Short-Cut" method and 6 points less than the raw scores for the Linear Method and the Traditional Item Statistics Method. All of the raw scores are lower than the raw score under the traditional standardization procedure.

Discussion

All of the equating methods which were used were able to answer the question about the ability of the current examinees. Their performance is better on the 46 common items than the performance of the reference examinees. Only the Rasch procedure, though, is able to answer the question about the items getting more difficult. The answer to this question is demonstrated in Figure 1.

The current and reference examinations were first calibrated individually. During this process the mean item difficulty is set equal to zero. For both item calibrations the mean ability estimates are quite close - .72 for the current examinees and .73 for the reference examinees. Thus, the two examinations are targeted at the same point in the two populations. When the link items are used to get the translation constant for the current examinees, it becomes apparent that the current examinees are more able than the reference examinees on the common items. Since they are more able, the

items on the current examination must be more difficult in order for the ability estimates to be so close when the mean item difficulty is set to zero.

Because the Rasch procedure is able to answer questions both about the items and the examinees, it seems to be the equating method which provides the most information. The answers it gives to the questions can help the standard setting body for physician's assistants decide whether they should equate their examinations from one year to the next. It might be that they want the standard to be higher as the training programs improve and the students gain more knowledge. The important thing is to learn facts about the examination and the population from the equating procedure so that a conscious decision can be made about the standard and one knows for sure whether one is giving the same or different scores to two groups.

A COMPARISON OF THREE EQUATING PROCEDURES ON
THE CERTIFYING EXAMINATION FOR PRIMARY CARE PHYSICIAN'S ASSISTANTS

Anita I. Bell
National Board of Medical Examiners

TABLE 1

Raw Scores for the Current Physician's Assistants Examination Scored Traditionally and Equated to a Reference Examination Via Four Methods

Standard Scores	Traditional Standardization	Reference-Current Link Items			
		Rasch	Linear	Linear Short-Cut	Traditional Item Statistics
700	185	175	182	184	179
650	175	167	172	173	169
600	165	158	161	162	159
550	155	147	151	151	149
500	146	137	140	140	139
450	136	125	130	129	129
420	130	118	124	122	124
410	128	116	122	120	122
400	126	114	119	118	120
350	116	102	109	107	110
300	107	91	98	96	100

A COMPARISON OF THREE EQUATING PROCEDURES ON
THE CERTIFYING EXAMINATION FOR PRIMARY CARE PHYSICIAN'S ASSISTANTS

Anita I. Bell
National Board of Medical Examiners

FIGURE 1